DOCUMENT RESUME

ED 053 904 RE 003 838

AUTHOR	Furlong, William
TITLE	Evaluator's Profile and Selection Procedures.
INSTITUTION	Indiana Univ., Bloomington. Measurement and Evaluation Center in Reading Education.
SPONS AGENCY	Office of Education (DHEW), Washington, D.C. Bureau of Research.
PUB DATE	70
NOTE	30p.
AVAILABLE FROM	Reading Program, School of Education, 202 Pine Hall, Indiana University, Bloomington, Ind. 47401 ($0.75)
EDRS PRICE	EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS	*Evaluation Methods, Evaluation Techniques, *Federal Aid, *Federal Programs, *National Surveys, *Profile Evaluation, *Program Evaluation, Questionnaires, Reading Programs, Reading Tests

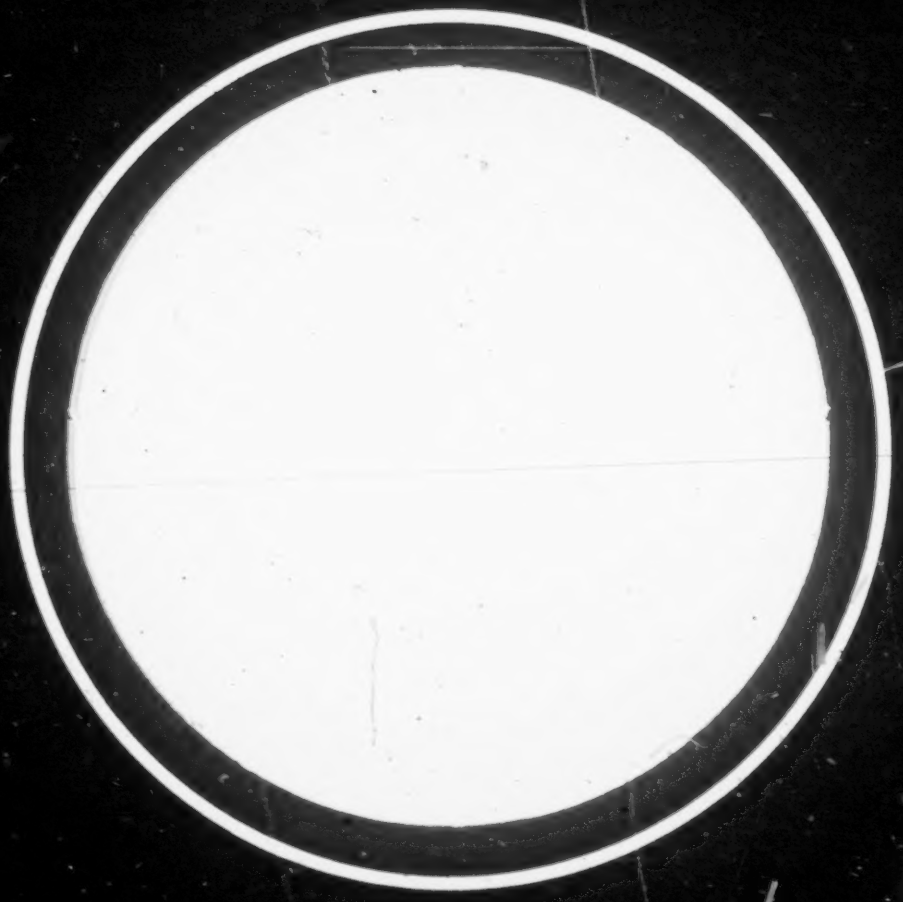
ABSTRACT

A profile of Title I and III reading program evaluators from data collected on a survey of 299 evaluators was made. The survey consisted of four sections: (1) The Evaluator--information on his educational background and training and on the programs with which he was connected; (2) Data Collection--information on the sorts of data the evaluator collected; (3) Reporting--information on the nature of the reports made and for whom they were intended; and (4) Tests--information on the types of tests used in the evaluations. Surprisingly few (30.4 percent) of the evaluators had had any formal course work in evaluation, yet 79.3 percent were responsible for writing evaluation reports, and the majority were responsible for evaluating objectives of projects, instructional procedures, materials, and staff. The majority of the evaluators developed criteria for evaluation from information from local teachers and consultants. It was questionable whether the teachers were actually evaluated. Most evaluation data came from examination of testing programs which are described in detail. Most reports were to state departments of education and most were written only to meet federal regulations. Tables of data, the survey, and survey responses are included. (AL)

OCCASIONAL PAPERS IN READING

INDIANA UNIVERSITY SCHOOL OF EDUCATION

ED0 53904



Evaluators' Profile

by

William Furlong

U. S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

Evaluators' Profile and Selection Procedures

by

William Furlong

Measurement and Evaluation Center in Reading Education
Indiana University, 1970

This study was supported in part by a grant
from U.S.O.E., H.E.W., Bureau of Research,
Contract #OEG-0-9-090473-4231 (010)

Analysis of Selected Items from Survey Form

The attached data is derived from 299 Survey Forms which were completed by evaluators of Reading Programs. Some information is extremely limited in that only a small percentage of evaluators actually responded to the item.

Items 8, 10, 11, 13, 15, and 16 (pages 4, 5, and 6 of questionnaire) received only a small percentage of responses. The apparent reason for this is that these questions were asked contingent upon preceding answers. If a prior answer was "yes," then the next question was applicable. The total response was thus reduced in proportion to the percentage of the preceding response pattern. Various comparisons can be developed through a careful analysis of all given data. The purpose of this summary is to illustrate some highlights which are seen within and between various sections of this survey.

I. The Evaluator

This section was designed to gather information about the preparation of the individuals involved in evaluation and some basic information about the programs with which they were connected. The title of the actual position varied from Assistant Superintendent to Evaluator to Teacher. Obviously, the title alone implied responsibilities beyond the scope of evaluation. In short, many individuals wore many hats. In spite of this, 237 (79.3%) respondents (N. 299) indicated that they were responsible for writing an evaluation report. This large percentage must be considered when question 4, page 2, is reached. Only 91 (30.4% evaluators, N. 299) actually took course work entitled "Evaluation" in a formal academic program. The significance of this is further developed when questions 7, 9, 12, and 14 (pages 4 and 5) in the data collection section are reached. Two-hundred twenty-five (75.3%) respondents were requested to evaluate the objectives of their projects; 144 (48.2%) were requested to evaluate instructional procedures; 154, (51.5%) participated in the evaluation of materials and 153 (51.2%) were involved in the evaluation of Title I Staff. There is an obvious dichotomy between formal training and occupational requirements in this one aspect of the evaluators' profile.

II. Data Collection

One indicator of the evaluator's perception of his effectiveness might be found in question 11 (page 5) in the data collection section. The response was asked contingent upon a "yes" answer to the preceding question. The percent was calculated on the total sample (N. 299) Thus, (N121), 40.5% of the evaluators made recommendations for change and 111 felt that change was made due to their suggestion. This is 37.7% of the total (N. 299).

With respect to data collection for evaluation purposes, several items are especially noteworthy. First, in developing criteria for evaluation

materials, local teachers (14.7%) and "experts" (consultants) (12.4%) were the primary source. Other Title I projects were noticeably low in preference for developing criteria. Second, teachers were evaluated primarily by locally devised scales (33.8%). The second choice response to this item was none (20.1%). Standard forms of evaluation were used by 7% of the respondents. Of course, 117 (39.1%) respondents did not reply to this question. Thus, the question arises: Were teachers evaluated? Third, the primary method of collecting data for total evaluation was an examination of the testing program (44.1%). (This is to be considered again in question 24, p. 9, where 263 - N. 299 - respondents -88% - indicated pre-test/post-test as their test design.) The second method was written reports from teachers (23.4%) and the third choice was discussion with teachers (16.1%).

III. Reporting

The recipient of the evaluator's report and information is another indication of the nature of the evaluator's role. The State Department of Education (42.5%), the Federal Government (16.4%) and the local school board (22.4%) are the recipients of the evaluation reports. Significantly, pupils (1.7%), parents (.7%), community groups (.3%), and the mass media (.3%) receive very little information concerning the results of Title I projects.

This is not so surprising when question 20 (page 7) is considered. The purpose of the report (in the eyes of the respondents) was, primarily, to meet Federal Regulations (44.1%) Project status (26.8%) and future financial support (17.4) were also noted as being especially important.

IV. Tests

As was noted earlier, the dominant source of evaluative information regarding the reading projects was test data. The least reliable measures of this progress were, in question 22, page 8, Reports from supervisors (3.7%), Pupils reading at grade level (4.7%), Parent response (.3%) and Classroom visitations (1.0%). Standardized tests (in question 23, page 8) were judged to be the most reliable measures of pupil progress. Academic grades on report cards were also considered to be reliable (32.8%) while reading grades were considered to be very unreliable (2%).

A certain degree of confusion about tests is reflected in the response pattern to question 25, page 9. The standardized test score means various things to various evaluators. Perhaps a common agreement on the definition of these terms might offer a stronger base for comparison of other responses.

Familiarity with standardized tests for specific purposes is suggested in the responses to qeustion 26, page 10. The Doren Diagnostic was selected as the best choice for both estimating Phonics-Word Attack Skills (32.4%) and Critical Reading Skills (14.4%). The Spitzer Test was

selected as best for estimating study skills (30.4%). The California Reading Test was selected for Paragraph Comprehension-Silent Reading (42.8%) The Dolch Test of 220 words was selected for sight vocabulary estimates (56.2%). Finally, the Nelson-Denny test was selected for secondary level reading comprehension (32.1%).

The above observations are only a few of the many insights which this survey offers into the profile of reading program evaluators. Many inferences can be drawn from the data as it is assembled in this survey. The primary observations concerning this information is that it is descriptive only. The sample indicates that 25% of the respondents had a degree beyond the Master's level, 51.5% had three or fewer hours in statistics, 65.6% had three or fewer hours in research design and 68.2% had no courses in evaluation. The data in no way reflects the success or failure of the evaluator in his task. This composite only reflects the educational background of educators involved in reading program evaluation. Also, the other data offered only gives a profile of perceptions and practices currently found in evaluation of reading programs.

COMPARISONS OF SELECTED RESPONSES

The following eight matrices are intended to illustrate some of the factors which are basic to the evaluators' profile. Essentially, there are two sets of comparisons. The first set (I, II, III, IV), relates the size of the evaluator's district to the degree he has earned, and to his study of evaluation, research design and statistics. The second set (V, VI, VII, VIII) relates the same four items to his test score.

To facilitate reading, it is useful to note that each cell contains three figures. The whole number, (in the upper left half), represents the number of evaluators responding to the particular cell. The decimal in the right triangle represents a percent figure of respondents. These figures are to be read across the matrix (right to left). The decimal in the bottom triangle represents the per cent of responses to the column. These figures are to be read down the column.

I

DEGREE/SIZE

Student Population	A.B.	M.A.	Spec.	Ph.D./Ed.D.	Degree earned by respondent
300	0 .0 .0	0 .0 .0	0 .0 .0	1 100. 2.2	
301-3000	3 10. 18.8	22 73.3 12.6	2 6.7 11.1	3 10. 6.5	
3001-12,000	5 6.5 31.3	61 79.2 35.1	5 6.5 27.8	6 7.8 13.0	
12,001-100,000	5 3.8 31.3	87 66.4 50.0	11 8.4 61.1	28 21.4 60.9	
100,000+	3 20.0 18.8	4 26.7 2.3	0 0.0 0.0	8 53.3 17.4	

This matrix includes comparisons of the degree earned by respondents to the questionnaire and the size of the district from which they are reporting.

KEY: Number responding Column % (Read Down) Row % (Read Across)

EXAMPLE: Read down the 4th column entitled "Ph.D/Ed.D":
Of the respondents from school districts with a student popluation of 300 or less, 2.2% had Ph.D.'s or Ed.D.'s.
Of the respondents from school districts with a student population of 301-3000, 6.5% had Ph.D.'s or Ed.D.'s.
Of the respondents from school districts with a student population of 3001-12,000, 13.0% had Ph.D.'s or Ed.D.'s.
Of the respondents from school districts with a student population of 12,001-100,000, 60.9% had Ph.D.'s or Ed.D.'s.
Of the respondents from school districts with a student population of 100,000+, 17.4% had Ph.D.'s or Ed.D.'s.

Read the bottom cell in the A.B. degree column with a size of over 100,000 as:
Of the respondents from districts of 100,000+, three had A.B. degrees. This is equal to 20% of all the respondents from districts of 100,000+.

II

HOURS IN STATISTICS/SIZE

Student Population	0	3	6	9	9+	Hours in statistics of respondent
300	0 0.0 0.0	0 0.0 0.0	1 100.0 1.6	0 0.0 0.0	0 0.0 0.0	
301-3000	9 30.0 15.3	9 30.0 12.2	7 23.3 11.3	5 16.7 17.2	0 0.0 0.0	
3001-12,000	17 22.4 28.8	30 39.5 40.5	20 26.3 32.3	8 10.5 27.6	1 1.3 3.8	
12,001 100,000	31 24.2 52.5	33 25.8 44.6	31 24.2 50.0	14 10.9 48.3	19 14.8 73.1	
100,000+	2 13.3 3.4	2 13.3 2.7	3 20.0 4.8	2 13.3 6.9	6 40.0 23.1	

This matrix is a cross reference of the education of respondents in statistics and the size of the district in which they work.

KEY: Number respondin[g] Column % (Read Down) Row % (Read Across)

Example: Read the bottom row across- Of the Districts with enrollment of 100,00 or more students:

13.3% had no hours in statistics
13.3% had 3 hours in statistics
20.0% had 6 hours in statistics
13.3% had 9 hours in statistics
40.0% had more than 9 hours in statistics

The total number of respondents from this size distract was (2, 2, 3, 2, 6) 15.

Read down the column marked 9+ - Of the respondents who had more than 9 hours in statistics,

0.0% were from districts with less than 300 enrollment
0.0% " " " " 301-3000 "
3.8% " " " " 3001-12,000 "
73.1% " " " " 12,001-100,000 "
23.1% " " " " more than 100,000 enrollment

The total number of respondents having more than 9 hours in statistics was (1, 19, 6) 26.

HOURS IN RESEARCH DESIGN/SIZE

Student Population	0	3	6	9	9+
	Hrs. in Research Design				
300	0 0.0 0.0	1 100.0 1.3	0 0.0 0.0	0 0.0 0.0	0 0.0 0.0
301-3000	12 37.9 12.4	8 27.6 10.3	7 24.1 13.2	1 3.4 8.3	2 6.9 11.8
3001-12,000	27 36.0 30.3	34 45.3 43.6	9 12.0 17.0	3 4.0 25.0	2 2.7 11.8
12,001-100,000	46 35.7 51.7	32 24.8 41.0	34 26.4 64.2	7 5.4 58.3	10 7.8 58.8
100,000+	5 33.3 5.6	3 20.0 3.8	3 20.0 5.7	1 6.7 8.3	3 20.0 17.6

This matrix is a cross reference of the education of the respondents in research design and the size of the district in which they work.

KEY: number responding | Column % (Read Down) | Row % (Read Across)

Example: Read across the bottom row - Of the districts with enrollment of 100,000+,

33.3% had 0 hours of research design
20.0% had 3 hours of research design
20.0% had 6 hours of research design
6.7% had 9 hours of research design
20.0% had more than 9 hours of research design

The total number of respondents from districts of 100,000 or more was (5,3,3,1,3)15.
Read down the column marked 9+ hours-Of the respondents who had 9+ hrs. of Research Design,
0.0% were from districts with less than 300 enrollment.
11.8% were from districts with 301-3000 enrollment.
11.8% were from districts with 3001-12,000 enrollment.
58.8% were from districts with 12,001-100,000 enrollment.
17.6% were from districts with 100,000+ enrollment.

The total number of respondents having more than 9 hours of research design was (2,2,10,3) 17.

IV

ANY COURSE IN EVALUATION/SIZE

Student Population	YES	NO	Any course in evaluation?
300	1 100.0 1.3	0 0.0 0.0	
301-3000	8 25.8 10.4	23 74.2 13.1	
3001-12,000	23 30.3 29.9	53 69.7 30.3	
12,001-100,000	40 31.0 51.9	89 69.0 50.9	
100,000+	5 33.3 6.5	10 66.7 5.7	

This matrix is a cross reference between whether or not the respondents had any course in evaluation and the size of the district in which they work.

KEY: Number Responding | X | Column % (Read Down) | X | Row % (Read Across) | X |

EXAMPLE: Read across the bottom row - Of the respondents from districts of more than 100,000,

33.3% had course(s) in evaluation
66.7% did not have a course in evaluation

The total number of respondents in this row (i.e. from districts of 100,000+) was (5,10) 15.
Read down the column marked "NO":
0.0% of the respondents from districts with less than 300 did not have an eval. course
13.1% of the respondents from districts with 301-3000 did not have an eval. course.
30.3% of the respondents from districts with 3001-12,000 did not have an eval. course.
50.9% of the respondents from districts with 12,001-100,000 did not have an eval.cours
5.7% of the respondents from districts with 100,000+ did not have an eval. course.

The total number of respondents who did not have an evaluation course was (0, 23, 53, 89, 10) 175.

V

TEST SCORE/DEGREE

Number Right	A.B.	M.A.	SPEC.	Ph.D./Ed.D.
0-15	6 6.6 37.5	76 83.5 43.4	3 3.3 16.7	6 6.6 13.0
16-21	7 5.7 43.8	84 68.9 48.0	11 9.0 61.1	20 16.4 43.5
22-28	3 7.1 18.8	15 35.7 8.6	4 9.5 22.2	20 47.6 43.5

Degree of respondent

This matrix is a cross reference of the test score made by the respondents and the degree they have completed.

KEY: Number Responding Column % (Read Down) Row % (Read Across)

EXAMPLE: Read across the bottom row - Of the people who correctly answered 22-28 of the items,

7.1% had A.B. degrees (3 individuals)
35.7% had M.A. degrees (15 individuals)
9.5% were specialists (4 individuals)
47.6% had Ph.D. degrees or Ed.D. degrees (20 individuals)

Read down the column marked Ph.D./Ed.D.:

13.0% of the respondents who answered 0-15 of the items correctly had Ph.D.'s or Ed.D.'
43.5% of the respondents who answered 16-21 of the items correctly had Ph.D.'s or Ed.D.
43.5% of the respondents who answered 22-28 of the items correctly had Ph.D.'s or Ed.D.

The total number of respondents who had Ph.D.'s or Ed.D.'s was (6,20,20) 46.

VI

TEST SCORE/HOURS IN STATISTICS

Number Correct	0	3	6	9	9+	Hrs. in Statistics
0-15	28 31.8 46.7	32 36.4 43.2	20 22.7 32.3	6 6.8 20.7	2 2.3 7.7	
16-21	29 23.6 48.3	36 29.3 48.6	34 27.6 54.8	16 13.0 55.2	8 6.5 30.8	
22-28	3 7.5 5.0	6 15.0 8.1	8 20.0 12.9	7 17.5 24.1	16 40.0 61.5	

This matrix is a cross reference of the test score made by the respondents and their education in statistics.

KEY: Number Responding Column % (Read Down) Row % (Read Across)

EXAMPLE: Read across the bottom row - Of the people who correctly answered 22-28 of the items,

7.5% had no hours in statistics (3 individuals)
15.0% had 3 hours in statistics (6 individuals)
20.0% had 6 hours in statistics (8 individuals)
17.5% had 9 hours in statistics (7 individuals)
40.0% had more than 9 hours in statistics (16 individuals)

Read down the column marked 9+ hours: (St=Statistics)
7.7% of the respondents who answered 0-15 of the items correctly had 9+ hrs. of St.
30.8% of the respondents who answered 16-21 of the items correctly had 9+ hrs. of St.
61.5% of the respondents who answered 22-28 of the items correctly had 9+ hrs. of St.

The total number of respondents who had more than 9 hours of statistic was (2, 3, 16) 26.

VII

TEST SCORE/HOURS IN RESEARCH DESIGN

Number Right	0	3	6	9	9+	Hrs. in Research Design
0-15	36 41.4 40.0	24 27.6 30.8	21 24.1 39.6	2 2.3 16.7	4 4.6 23.5	
16-21	48 39.7 53.3	45 37.2 57.7	19 15.7 35.8	2 1.7 16.7	7 5.8 41.2	
22-28	6 14.3 6.7	9 21.4 11.5	13 31.0 24.5	8 19.0 66.7	6 14.3 35.3	

This matrix is a cross reference of the test score made by the respondents and their education in research design.

KEY: Number Responding [X] Column % (Read Down) [X] Row % (Read Across) [X]

EXAMPLE: Read across the bottom row - Of the people who correctly answered 22-28 of the items,

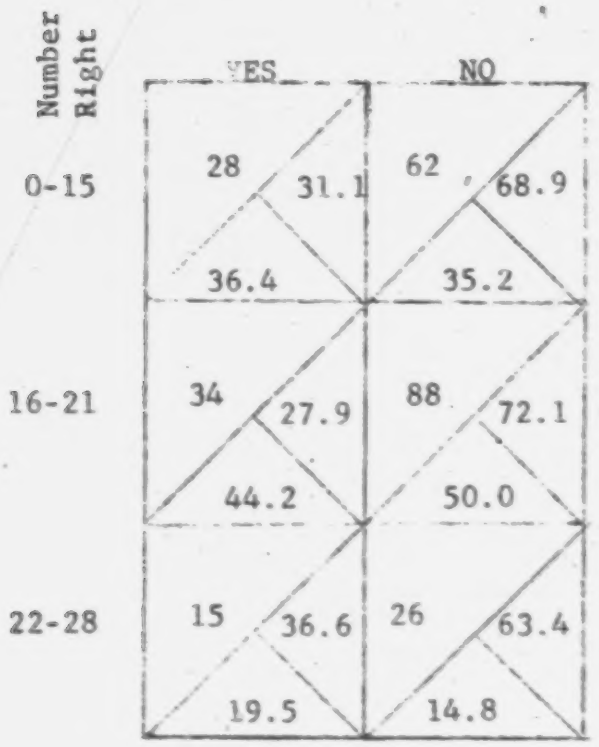
14.3% had no hours in research design.
21.4% had 3 hours in research design.
31.0% had 6 hours in research design.
19.0% had 9 hours in research design.
14.3% had more than 9 hours in research design.

The total number of respondents who made correct responses to 22-28 of the items was (6, 9, 13, 8, 6) 42.
Read down the column marked 9+ hours: (R.D. = Research Design)
23.5% of the respondents who answered 0-15 of the items correctly had 9+ hrs. in R.D.
41.2% of the respondents who answered 16-21 of the items correctly had 9+ hrs. in R.D.
35.3% of the respondents who answered 22-28 of the items correctly had 9+ hrs. in R.D.

The total number of respondents who had more than 9 hours in research design was (4, 7, 6) 17.

VIII

TEST SCORE/ANY COURSE IN EVALUATION



This matrix is a cross reference of the test score made by the respondents and whether or not they had any course(s) in evaluation.

KEY: number responding: Column % (Read Down) Row % (Read Across)

EXAMPLE: Read across bottom row - Of the respondents who answered 22-28 of the items correctly,

36.6% had course(s) in evaluation (15 individuals)
63.4% did not have a course in evaluation (26 individuals)

Read down the column marked "NO": (E.C.=Evaluation Course)

35.2% of the respondents who answered 0-15 of the items correctly did not have an E.C.

50.0% of the respondents who answered 16-21 of the items correctly did not have an E.C.

14.8% of the respondents who answered 22-28 of the items correctly did not have an E.C.

The total number of respondents who did not have an evaluation course was (62, 88, 26) 176.

U. S. O. E. Survey of Evaluators
Title I and III Readin Pro rams

Measurement and Evaluation
Center in Readin Education
Carl B. Smith, Director

School of Education
Indiana University
Bloomin ton, Indiana

SURVEY DATA

The following information is a compilation of data from individual survey forms. The questions are listed here in the same order as on the survey form. The frequency of response is followed by the percent of responses (in parentheses) for each question. Please note that some questions ask for a ranking response (1, 2, 3) and others require a simple multiple response (i.e. mark the tests used, etc.). A sample of the original survey form is included in the appendix to this report.

Page 1

G. Were you responsible for writing an evaluation report for your 1968-69 Title I or III reading project?

_____	yes	237	(79.3%)
_____	no	60	(20.1%)

Page 2

Educational Background

Highest degree completed is:

_____	Bachelor	20	(6.7%)
_____	Master	204	(68.2%)
_____	Specialist	21	(7.0%)
_____	Doctor	52	(17.4%)

Graduate hours in statistics are:

_____	none	71	(23.7%)
_____	3	83	(27.8%)
_____	6	74	(24.7%)
_____	9	35	(11.7%)
_____	more than 9	29	(9.7%)

Graduate hours in research design are:

_____	none	101	(33.8%)
_____	3	95	(31.8%)
_____	6	61	(20.4%)
_____	9	15	(5.0%)
_____	more than 9	19	(6.4%)

Have you taken any course work entitled "evaluation"?

_____	yes	91	(30.4%)
_____	no	204	(68.2%)

Page 3

1. Within the last 18 months I have been involved in the following programs for Title I and/or III evaluation purposes:
(Three choices. Respondents who started with "A" are included in column 1; those who started with "B" are in column 2, etc.)

	1	2	3
_____A. Inservice workshop	169 (56.5%)		
_____B. Extension course	2 (0.7%)	18 (6.0%)	
_____C. University consultantship	9 (3.0%)	29 (9.7%)	5 (1.7%)
_____D. Publishers workshop	1 (0.3%)	13 (4.3%)	13 (4.3%)
_____E. Program of local I.R.A. council	2 (0.7%)	15 (5.0%)	8 (2.7%)
_____F. Program of state department of public instruction	43 (14.4%)	45 (15.1%)	24 (8.0%)
_____G. Other (please specify)	31 (10.4%)	8 (2.7%)	8 (2.7%)

2. Number of years of operation of your Title I and/or III project is:

_____A. 1	14 (4.7%)
_____B. 2	12 (4.0%)
_____C. 3	93 (31.1%)
_____D. 4	173 (57.9%)

3. The total number of public school pupils (K-12) in your district is:

_____A. 300 or less	1 (0.3%)
_____B. 301-3,000	34 (11.4%)
_____C. 3,001-12,000	81 (27.1%)
_____D. 12,001-100,000	164 (54.8%)
_____E. 100,000+	17 (5.7%)

4. What process led to the establishment of the reading project?

_____A. A survey of needs resulting in a project plan	186 (62.2%)
_____B. A project plan developed by a committee of teachers & administrators	50 (16.7%)
_____C. A plan developed by administrators and supervisors	43 (14.4%)
_____D. A community group request for a project	3 (1.0%)
_____E. Do not know or not sure	10 (3.3%)

Page 4

5. Reading instruction for this project is typically conducted in groups of:
(check most appropriate term)

_____A. 2-3	29 (9.7%)
_____B. 3-10	172 (57.5%)
_____C. 10-15	55 (18.4%)
_____D. Whole class	11 (3.7%)
_____E. Individualized instruction	20 (6.7%)

6. This method of grouping was prescribed because:

_____	A. More children are reached	35 (11.7%)
_____	B. It is suited to the size of the available classroom	14 (4.7%)
_____	C. Pupil needs are met best	236 (78.9%)
_____	D. It is less expensive	3 (1.0%)

II--Data Collection

7. As an evaluator of the reading project, were you requested to evaluate the objectives of the project?

_____	Yes	225 (75.3%)
_____	No	61 (20.4%)

8. If "Yes" above, what procedures did you use in making your evaluation of the objectives? (rank 1, 2, 3).

		1	2	3
_____	A. Discussed objectives with local specialists to see whether they were feasible	25(8.4%)	73(24.4%)	4(1.3%)
_____	B. Related objectives to surveyed needs	85(28.4%)	21(7.0%)	6(2.0%)
_____	C. Submitted objectives to a board of experts	1(0.3%)	3(1.0%)	31(10.4%)
_____	D. Other (Specify) _____	9(3.0%)	4(1.3%)	21(7.0%)

9. Were you requested to evaluate instructional procedures used in the project?

_____	Yes	144 (48.2%)
_____	No	142 (47.5%)

Page 5

10. If "Yes" above, did you make recommendations to change any instructional procedures?

_____	Yes	121 (40.5%)
_____	No	37 (12.4%)

11. If "Yes" above, were changes made during the project as a result of your recommendations?

_____	Yes	111 (37.1%)
_____	No	13 (4.3%)

12. Did you participate in the evaluation of teaching materials?

_____	Yes	154 (51.5%)
_____	No	131 (43.8%)

13. If "Yes", indicate the sources for developing the criteria for evaluating the materials. (Rank 1, 2, 3.)

		1	2	3
_____A.	Books on evaluation	6(2.0%)	8(2.7%)	14(4.7%)
_____B.	Expert (e.g. consultant)	37(12.4%)	34(11.4%)	8(2.7%)
_____C.	Publishers representatives	3(1.0%)	9(3.0%)	25(8.4%)
_____D.	Local teachers	44(14.7%)	32(10.7%)	14(4.7%)
_____E.	Other Title I projects	7(2.3%)	12(4.0%)	22(7.4%)

14. Were you involved in evaluation of Title I staff?

_____	Yes	153 (51.2%)
_____	No	134 (44.8%)

15. If "Yes" above, indicate methods used to gather data. (Rank 1, 2, 3.)

		1	2	3
_____A.	Personal classroom observation	72(24.1%)	16(5.4%)	6(2.0%)
_____B.	Teacher's log or lesson plan	2(0.7%)	18(6.0%)	21(7.0%)
_____C.	Principal report	12(4.0%)	34(11.4%)	13(4.3%)
_____D.	Questionnaire	8(2.7%)	12(4.0%)	14(4.7%)
_____E.	Interview	1(0.3%)	10(3.3%)	14(4.7%)
_____F.	Consultant's report from observation	3(1.0%)	4(1.3%)	7(2.3%)
_____G.	Supervisor's report	1(0.3%)	5(1.7%)	14(4.7%)
_____H.	Other (specify)	3(1.0%)	1(0.3%)	2(0.7%)

Page 6

16. Indicate the teacher evaluation instrument used. (Two choices. Respondents who started with "A" are included in Column 1; those who started with "C" are in Column 2.)

		1	2
_____A.	Flanders Interaction Analysis	3(1.0%)	
_____B.	Amidon Teacher Analysis Form	0(0.0%)	0(0.0%)
_____C.	Locally devised rating scale	101(33.0%)	2(0.7%)
_____D.	None	60(20.1%)	0(0.0%)
_____E.	Other (specify)	18(6.0%)	10(3.3%)

17. Indicate methods used to collect data for your total evaluation. (Rank 1, 2, 3.)

		1	2	3
_____A.	Visit reading classes	52(17.4%)	29(9.7%)	29(9.7%)
_____B.	Discussions with teachers	14(4.7%)	57(19.1%)	48(16.1%)
_____C.	Examine results of testing program	132(44.1%)	41(13.7%)	30(10.0%)
_____D.	Pupil reactions	5(1.7%)	14(4.7%)	44(14.7%)
_____E.	Written reports from teachers	18(6.0%)	70(23.4%)	29(9.7%)
_____F.	Reports from supervisors	8(2.7%)	17(5.7%)	35(11.7%)
_____G.	Others (Describe)	5(1.7%)	4(1.3%)	12(4.0%)

Page 7

III--Reporting

18. To whom are the results of Title I projects reported? (Rank 1, 2, 3.)

		1	2	3
_____A.	U.S. Government	3(1.0%)	1(0.3%)	5(1.7%)
_____B.	State department of education	49(16.4%)	30(10.0%)	16(5.4%)
_____C.	Local school board	127(42.5%)	62(20.7%)	15(5.0%)
_____D.	Community groups (specify)	31(10.4%)	67(22.4%)	67(22.4%)
_____E.	Newspaper, radio, etc.	1(0.3%)	0(0.0%)	16(5.4%)
_____F.	Parents (specify)	1(0.3%)	2(0.7%)	5(1.7%)
_____G.	Faculty and staff	2(0.7%)	10(3.3%)	13(4.3%)
_____H.	Pupils	17(5.7%)	30(10.0%)	66(22.1%)
_____I.	Supervisory personnel	5(1.7%)	1(0.3%)	1(0.3%)
_____J.	Other (specify)	11(3.7%)	41(13.7%)	32(10.7%)

19. Indicate the type(s) of reports you were requested to make. (Three choices. Respondents who started with "A" are included in Column 1; those who started with "B" in Column 2, etc.)

		1	2	3
_____A.	Anecdotal	79(26.4%)		
_____B.	Narrative	162(54.2%)	68(22.7%)	
_____C.	Statistical	44(14.7%)	157(52.5%)	60(20.1%)
_____D.	Other (specify)	6(2.0%)	6(2.0%)	8(2.7%)

20. For what purposes are reports of Title I projects made? (Rank 1, 2, 3).

		1	2	3
_____A.	Future financial support	20(6.7%)	52(17.4%)	52(17.4%)
_____B.	Local publicity	1(0.3%)	1(0.3%)	9(3.0%)
_____C.	Federal regulation	132(44.1%)	35(11.7%)	35(11.7%)
_____D.	Decision regarding status of project	54(18.1%)	80(26.8%)	49(16.4%)
_____E.	Periodic progress during the year	32(10.7%)	34(11.4%)	33(11.0%)
_____F.	Annual report to Board of Education	5(1.7%)	39(13.0%)	52(17.4%)

21. What means were used to make public the results of the Title I project? (Three choices. Respondents who started with "A" are included in Column 1; those who started with "B" are in Column 2, etc.)

		1	2	3
_____A.	Newspaper articles	172(57.5%)		
_____B.	Reports to parents, written	32(10.7%)	72(24.1%)	
_____C.	Radio interviews	0(0.0%)	14(4.7%)	12(4.0%)
_____D.	Reports to P.T....	17(5.7%)	48(16.1%)	36(12.0%)
_____E.	Reports to community service organizations	4(1.3%)	21(7.0%)	45(15.1%)
_____F.	Reports to Board of Education	38(12.7%)	54(18.1%)	53(17.7%)
_____G.	Written report--circulated in community	1(0.3%)	4(1.3%)	17(5.7%)
_____H.	Other (specify)	11(3.7%)	12(4.0%)	4(1.3%)

Page 8

IV--Tests

22. In your judgment, the most reliable measures of the degree of success of the reading project are: (Rank 1, 2, 3.)

		1	2	3
_____A.	Statistics derived from pre- and post-tests	108(36.1%)	46(15.4%)	56(18.7%)
_____B.	Pupils reading at grade level	14(4.7%)	20(6.7%)	14(4.7%)
_____C.	Parent response	1(0.3%)	6(2.0%)	19(6.4%)
_____D.	Responses from classroom teachers	33(11.0%)	62(20.7%)	56(18.7%)
_____E.	Extent of pupil participation in activities	34(11.4%)	28(9.4%)	26(8.7%)
_____F.	Pupils reading near expectancy level	64(21.4%)	57(19.1%)	30(10.0%)
_____G.	Reading teacher reports	19(6.4%)	40(13.4%)	41(13.7%)
_____H.	Reports from supervisors	3(1.0%)	7(2.3%)	11(3.7%)
_____I.	Classroom visitation	3(1.0%)	8(2.7%)	16(5.4%)

23. In your judgment, the most reliable measures of pupil progress are: (Rank 1, 2, 3).

		1	2	3
_____A.	Standardized test results	1(0.3%)	0(0.0%)	1(0.3%)
_____B.	Reading grades on report card	120(40.1%)	50(16.7%)	54(18.1%)
_____C.	Academic grades on report card	0(0.0%)	4(1.3%)	6(2.0%)
_____D.	Reading teacher evaluation	0(0.0%)	6(2.0%)	8(2.7%)
_____E.	Oral reading in the classroom	98(32.8%)	94(31.4%)	25(8.4%)
_____F.	Pupil's self-evaluation	6(2.0%)	12(4.0%)	13(4.3%)
_____G.	Classroom teacher evaluation	21(7.0%)	35(11.7%)	55(18.4%)
_____H.	Informal test results	19(6.4%)	45(15.1%)	60(20.1%)
_____I.	Number of books read by pupil	18(6.0%)	28(9.4%)	34(11.4%)
_____J.	Completion of basal reader	1(0.3%)	6(2.0%)	12(4.0%)
_____K.	Completion of reading workbook	0(0.0%)	1(0.3%)	1(0.3%)

Page 9

24. What test design was used in evaluating the project?

_____A.	Pre-test/post-test	263(88.0%)
_____B.	Local data comparison	12(4.0%)
_____C.	National norm comparison	8(2.7%)
_____D.	Test of difference between groups	3(1.0%)
_____E.	Other (describe)	4(1.3%)

25. Standardized reading test scores indicate:

_____A.	Independent reading level	88(29.4%)
_____B.	Instructional reading level	99(33.1%)
_____C.	Frustration reading level	81(27.1%)

Use the following list of tests to answer the question below:

1. Doren Diagnostic
2. Spitzer Test
3. California Reading Test
4. Dolch Test of 220 Words
5. Nelson-Denny
6. Gray Oral
7. Informal Reading Inventory
8. Peabody Picture Vocabulary Test

Page 10

26. Which of the following reading skills are tested by the instruments indicated above? Write the number of the test next to the skill that it purports to measure.

_____A. Phonics-word attack

1	2	3	4	5	6	7	8
97(32.4%)	3(1.0%)	9(3.0%)	19(6.4%)	2(0.7%)	30(10.0%)	20(6.7%)	7(2.3%)

_____B. Sight vocabulary

1	2	3	4	5	6	7	8
10(3.3%)	1(0.3%)	7(2.3%)	168(56.2%)	3(1.0%)	9(3.0%)	9(3.0%)	21(7.0%)

_____C. Study skills

1	2	3	4	5	6	7	8
12(4.0%)	91(30.4%)	40(13.4%)	1(0.3%)	10(3.3%)	2(0.7%)	15(5.0%)	0(0.0%)

_____D. Paragraph comprehension-silent reading

1	2	3	4	5	6	7	8
4(1.3%)	3(1.0%)	128(42.8%)	4(1.3%)	23(7.7%)	18(6.0%)	43(14.4%)	0(0.0%)

_____E. Critical reading skills

1	2	3	4	5	6	7	8
43(14.4%)	9(3.0%)	38(12.7%)	4(1.3%)	18(6.0%)	26(8.7%)	23(7.7%)	0(0.0%)

_____F. Secondary level reading comprehension

1	2	3	4	5	6	7	8
2(0.7%)	11(3.7%)	53(17.7%)	5(1.7%)	102(34.1%)	8(2.7%)	10(3.3%)	0(0.0%)

The following are types of tests often used in reading projects:

1. Achievement test (standardized, California)
2. Diagnostic reading test (standardized, group, Bond-Clymer-Hoyt)
3. Individual diagnostic reading test (Durrell)
4. Informal reading inventory
5. Interest inventory
6. Language and concept inventory (Peabody)

27. Use the above listed types of tests, and write the number of the type which would be appropriate for measuring or evaluating the following objectives:

_____A. Improve all pupils reading through a developmental program

1	2	3	4	5	6
153(51.2%)	72(24.1%)	11(3.7%)	10(3.3%)	2(0.7%)	4(1.3%)

_____B. Correct specific reading difficulties through a remedial reading program

1	2	3	4	5	6
6(2.0%)	58(19.4%)	175(58.5%)	16(5.4%)	0(0.0%)	3(1.0%)

_____C. Enrich children's experiences for reading through field trips

1	2	3	4	5	6
3(1.0%)	0(0.0%)	0(0.0%)	29(9.7%)	124(41.5%)	82(27.4%)

_____D. Encourage wide reading through appropriate book selection

1	2	3	4	5	6
6(2.0%)	4(1.3%)	2(0.7%)	89(29.8%)	126(42.1%)	15(5.0%)

V--Measurement

The following items were taken from a test constructed by Samuel T. Mayo, Loyola University, Chicago, Illinois, U.S.O.E. Project No. 5-0807/Contract No. OE4-10-001: Pre-Service Preparation of Teachers in Educational Measurement.

DIRECTIONS

Please circle the answer that you think is most appropriate. Do not consult other sources. Make your selection on the basis of your present knowledge.

1. The essential difference between standardized and unstandardized tests lies in

    1. their validity
    2. their objectivity
    3. the availability of norms
    4. the discriminatory capacity of their items

2. Advocates of "culture fair" tests of mental ability can most justifiably criticize the Stanford-Binet because of its emphasis in measuring

    1. organization of ideas
    2. fluency of ideas
    3. verbal abilities
    4. innate abilities

3. If a student wanted to find the most appropriate achievment test in arithmetic, he should consult

    1. publishers' catalogues
    2. Buros' Mental Measurements Yearbook
    3. Journal of Experimental Education
    4. the most recent texts in the teaching of arithmetic

4. If a teacher wanted to determine how well a standardized test would measure the objectives which she had been trying to teach, it would be best for her to examine

    1. the test itself
    2. critical reviews of the test
    3. the manual for the test
    4. recent studies in which the test had been used

5. It is more appropriate to discuss the mental stanine of a child with a parent than the child's I.Q. because

    1. the stanine is a more valid measure of intelligence
    2. the I.Q. appears more precise than it actually is
    3. mental stanines are more highly correlated with achievement
    4. parents are better kept in doubt with reference to the child's ability

6. Which of the following types of items is well adapted to evaluating student knowledge of numerous technical terms?

   1. true-false
   2. multiple choice
   3. matching
   4. analogy

7. The term objective, when used to label an education test, describes

   1. a characteristic of the scoring process
   2. a typographic feature of the test
   3. the degree of standardization of the test
   4. the content limitations of the questions

8. Sue answered correctly 25 out of 50 items on an arithmetic test. What interpretation can be made of Sue's performance on the test?

   1. Sue placed at the 50th percentile
   2. Sue needs remedial work in arithmetic
   3. Sue knows about one-half of the material in arithmetic taught in her grade
   4. No interpretation of the score is possible on the basis of the information given

9. Problems arise in attempting to develop measures of ultimate goals mainly because

   1. measurement methods have not given proper weight to all goals
   2. teachers have been reluctant to depart from traditional testing methods
   3. group norms with which to compare results are not available
   4. such goals concern behaviour not usually observable under classroom conditions

10. Which of the following is an <u>untrue</u> statement about instructional goals?

    1. The worth of a goal is determined by its measurability.
    2. A two-way chart helps to relate content to educational goals.
    3. One test can usually measure only a few goals.
    4. Content and method vary directly with goals.

11. Why should behavioural objectives as contrasted with content objectives <u>best</u> be restricted in number?

    1. to facilitate organization of a course
    2. to promote their operational definition
    3. to enable a teacher to keep them constantly in mind during instruction

12. "Washington, D. C., is the most important city in the United States." Why is this a poor true-false item?

    1. It is ambiguous.
    2. It is too easy.
    3. It is too brief.
    4. It is too factual.

13. "Philadelphia was the capitol and largest city in the United States for a number of years." Why is this a poor true-false item?

1. It is ambiguous.
2. It involves more than one idea.
3. It does not have a good answer.
4. It is too long.

14. Validity is determined by finding the correlation between scores on

1. the even numbered items on a test and the odd numbered items on that
2. one form of a test and another form of that same test test
3. a test and some independent criterion
4. two administrations of the same test

15. For determining reliability, for retesting doubtful cases, or for measuring growth, it is most useful to have

1. equivalent forms
2. adequate norms
3. objectivity and interpretability
4. logical and empirical validity

16. If the reliability of an arithmetic test is .50, and if the length is doubled, the reliability would

1. increase
2. decrease
3. remain the same
4. change in some indeterminate way

17. Upon receiving intelligence test scores for her class, a teacher is surprised to learn that a pupil she has always considered as "average" has an I.Q. of 84. Of the following, what is her most appropriate course of action?

1. Check the pupil's cumulative record for the results of previously administered achievement and intelligence tests.
2. Evaluate her attitude toward the pupil's performance in class to learn whether she has been grading him too leniently.
3. Discuss the test results with the pupil to learn whether he was ill on the day of the test.
4. Recognize that the pupil is achieving far beyond his capacity and encourage him to continue.

18. What is the chief obstacle to effective homogeneous grouping of pupils on the basis of their educational ability?

1. resistance of children and parents to discriminations on the basis of ability
2. difficulty of developing suitably different teaching techniques for the various levels
3. increased costs of instruction as the number of groups increases and their average size decreases
4. wide differences in the level of development of various abilities within individual pupils

19. A diagnostic test which provides the teacher with a profile of scores is of <u>little</u> value unless

1. the sub-tests which make up the profile are quite reliable
2. the test has reliable norms
3. the test has been shown to be a valid predictor of future achievement
4. the scores are reported in terms of percentile ranks

20. In order to compute a correlation coefficient between traits A and B, it is necessary to have

1. measures of trait A on a group of persons, and of trait B on another
2. one group of persons, some who have both A and B, some with neither, and some with one but not the other
3. two groups of persons, one which could be classified as A or not A, the other as B or not B
4. measures of traits A and B on each person in one group

21. Test norms are most satisfactory when the sample of pupils or students used in establishing the norms

1. consists of nearly all pupils or students taking the test prior to the time the norms are published
2. is representative of a clearly defined population with which it is appropriate to make comparisons
3. ranges over all the grade levels in which the test is likely to be used
4. includes all schools volunteering to participate in the standardization testing

22. A good diagnostic test most differs from a good survey achievement test in

1. reliable and valid measurement of skills
2. identifying causes of weaknesses
3. possessing equivalent forms so that growth in achievement can be measured
4. identifying pupils whose achievement is unsatisfactory

23. The State of X has a state-wide testing program. As a basis for revising the objective examination in science, a set of papers from the top and bottom quarter of the total group tested was analyzed. The per cent passing each item was determined. <u>Other things being equal</u>, which of the following items would one be most likely to keep in the test?

1. top quarter--98%, bottom quarter--92%
2. top quarter--80%, bottom quarter--40%
3. top quarter--70%, bottom quarter--75%
4. top quarter--25%, bottom quarter--10%

24. A student scores 35 on a vocabulary test. The mean for the class is 37.3 and the standard deviation is 8.4. His z-score is

1. .27
2. .23
3. -.27
4. -.44

25. What does the percentile equivalent of a raw score indicate?

1. the percent of a group making scores above the midpoint of that raw score interval
2. the percent of a group making scores between the upper and lower limits of that raw score interval
3. the percent of a group making scores lower than the midpoint of that raw score interval
4. the percent of items of the test which must be answered correctly to get that raw score

26. In a particular situation the frequency distribution of scores on a standardized test is found to be approximately normal. This should be regarded as

1. common and highly desirable
2. common but not especially desirable
3. rare and highly desirable
4. rare and not especially desirable

27. If a certain test is taken by a group of high school seniors, and is found to correlate .62 with freshman grades received in college by these same seniors, one can say that

1. the test is a valid predictor of college aptitude
2. the test is not a reliable measure of college success
3. approximately two-thirds of those taking the test will be successful in college
4 students who score lower than 62 will be unsuccessful in college

28. The standard error of measurement is a numerical figure which indicates

1. the number of points a student's test score is in error in relation to the score he should make
2. the number of points the mean score for the test is in error
3. a range of scores within which the student's true score most probably falls
4. the reliability of the test norms

Evaluators' Test Response Profile

The twenty-eight items which were included as a survey of evaluators' knowledge about principles of evaluation are listed below. The item number, the correct answer and the percentages of each responses are included. The percent figure, which is underlined, is the correct choice. The actual questions are included in the appendix as the second half of the questionnaire.

		Percent of Responses (0=no answer)				
Item	Key	0	1	2	3	4
1	3	4%	20%	4%	<u>69%</u>	3%
2	3	5%	2%	7%	<u>76%</u>	9%
3	2	5%	7%	<u>75%</u>	5%	8%
4	2	5%	56%	<u>12%</u>	18%	10%
5	2	3%	5%	<u>74%</u>	18%	0%
6	3	4%	2%	33%	<u>44%</u>	17%
7	1	3%	<u>46%</u>	4%	12%	35%
8	4	1%	7%	0%	3%	<u>88%</u>
9	4	5%	26%	6%	7%	<u>55%</u>
10	1	3%	<u>70%</u>	8%	13%	6%
11	3	7%	18%	31%	<u>44%</u>	0%
12	1	1%	<u>84%</u>	7%	5%	3%
13	2	3%	15%	<u>78%</u>	4%	1%
14	3	4%	7%	17%	<u>64%</u>	9%
15	1	4%	<u>73%</u>	10%	8%	5%
16	1	4%	<u>31%</u>	10%	21%	33%
17	1	5%	<u>74%</u>	1%	3%	18%
18	4	2%	14%	5%	3%	<u>75%</u>
19	1	2%	<u>54%</u>	17%	14%	13%
20	4	5%	6%	18%	9%	<u>61%</u>
21	2	2%	5%	<u>82%</u>	8%	2%
22	2	3%	14%	<u>74%</u>	5%	5%
23	2	11%	6%	<u>51%</u>	13%	19%
24	3	20%	11%	7%	<u>58%</u>	4%
25	3	5%	4%	25%	<u>51%</u>	14%

Evaluators' Test Response Profile (Cont.)

Item	Key	Percent of Responses 0	1	2	3	4
26	2	6%	44%	28%	14%	7%
27	1	9%	54%	24%	12%	1%
28	3	12%	7%	10%	64%	7%

Respondents	N = 296
Items	N = 28.0
Means	17.0743
Sigmas	5.2480
Alphas	0.8199